

СЕМАНТИЧЕСКАЯ ПАУТИНА, ФАНТОМ, ИЛИ «ИНТЕРНЕТ БУДУЩЕГО»

Ефимов А.А.

ГОУ ВПО «Уральский государственный педагогический
университет», г. Екатеринбург

Заводя речь о концепции семантической паутины (Semantic Web) стоит сначала определить некоторые термины, разобраться в том, что же такое, эта семантическая паутина.

Говоря «Интернет», обычно, мы понимаем под этим словом некую застывшую в развитии технологическую единицу, тем самым сильно упрощая истинное положение дел. На самом деле — что, впрочем, тоже не секрет — «сеть» прошла два этапа своей эволюции, сделавших её такой, какой мы видим её сегодня. Первым этапом, уходящим корнями в 70-ые годы, был собственно Интернет — выросшая из военного проекта глобальная компьютерная сеть, узлы которой обменивались данными, не требовавшими обработки в реальном времени (письма, файлы и т.п.). В начале 90-ых родилась World Wide Web — Всемирная паутина, само название которой подразумевает упрощение и популяризацию технологий,

бывших уделом учёных, военных и компьютерных гениев. Web предоставил унифицированный способ доступа к различным типам данных. В результате этого, вместилищем информации стали не файловые архивы, а web-сайты, работа online сделалась проще, наглядней и интересней, а сама «сеть» превратилась из глобальной компьютерной в глобальную информационную. Но и Web, по сути, без особых изменений доживший до настоящего времени, не стал последним этапом: на очереди следующая эволюционная ступень, именуемая Semantic Web — сеть семантическая [<http://www.w3.org/2001/sw>].

Семантическая паутина была задумана консорциумом W3 [<http://www.w3.org/>] очень давно. С середины 90-х публиковались разные статьи и заметки, которые не привлекали особого внимания широкой общественности, но однажды, председателем консорциума W3 была опубликована статья, обратившая на себя внимание всего Интернет сообщества:

17 мая 2001 года в журнале Scientific American появилась статья Тима Бернерса-Ли (Tim Berners-Lee), Джеймса Хендлера (James Handler) и Оры Лассила (Ora Lassila) под названием «The Semantic WEB»². Надо отметить, что Тим Бернерс-Ли — это отец WWW. Он не только первым предложил концепцию всемирной паутины, но и воплотил её в жизнь. К словам такого человека просто нельзя не прислушиваться.

Стиль описания, напомнил мне громкие пророчества манифеста «Some Computer Science Issues in Ubiquitous Computing» [3], которые, кстати, сбываются, про то, что главной проблемой является не повсеместная компьютеризация жизни, а конфиденциальность личной жизни в таких условиях. Или опять же сюжеты про цифровые дома из японской анимационной фантастики — во время того, как ты сидишь на работе, электронная штучковина внутри твоего скромного жилища находит оставленное задание «приготовить компот», холодильник через сеть связывается с ближайшим гастрономом, вместе они решают, что для приготовления необходимы сухофрукты и вода, после этого прямо к холодильнику, с ближайшего рынка, т.к. это ровно на 35.987% выгоднее гастронома, доставляются отборные сухофрукты. Кухонный процессор

² Доступен перевод [http://ezolin.pisem.net/logic/semantic_web_rus.html], выполненный Евгением Золиным.

снова идет в сеть, ищет рецепт компота из сухофруктов, и его не ставит в тупик то обстоятельство, что рецептов несколько, он то знает не только предпочтения хозяина, но и его возраст, состояние здоровья и многое другое, поэтому находит такой рецепт компота разжижающего кровь «для тех, кому до шестнадцати и старше». Цифровой дом так же находит в сети информацию о том, что для подачи этого напитка стол должен быть сервирован стаканом, причем лучше высоким, с прямыми стенками...

Впечатляющий футуризм. У этой статьи, как и у всех подобных была только одна цель — привлечь внимание к Semantic Web всех, кого только можно было. Зачем — непонятно. Скорей всего, авторы просто сделали открытие (а Semantic Web на самом деле впечатляющее открытие), и, находясь, некоторое время в просветленном состоянии решили написать все, что думают.

В двух словах, под семантической паутиной понимается структура, наполненная информацией о смысловом содержании, понятной машинам. В самом деле, задумайтесь: ведь современные web-страницы созданы для человека, компьютерным программам не под силу проанализировать их содержимое — искусственный интеллект ведь пока не изобрели. Главной задачей семантической паутины станет добавление специальных данных, которые позволят компьютерным программам понимать web-страницы. Скромное на первый взгляд добавление выведет Интернет на новый виток эволюции. Представьте, что браузеры научатся «понимать», о чём говорится на web-страничках, которые они открывают. Это сразу же на порядок повысит эффективность работы с web-ресурсами: компьютеры сами, без помощи человека смогут находить нужную нам информацию, самостоятельно отсеивать сор, похожий словарно, но отличный по смыслу, будут в состоянии автоматически выполнять самые различные поручения по работе с web-контентом, вроде заказа мест в ресторане или подбора рефератов. Человеческий интеллект, который сегодня является необходимым инструментом при работе с Сетью (обратите внимание, как вы работаете, к примеру, с вашим любимым поисковиком), станет машинам не нужен.

На самом деле всё не так страшно и фантастично как кажется. Для того чтобы выделить имя или e-mail адрес средствами HTML были созданы тэги <NAME> и <ADRESS>, но на их стандартизацию ушли

годы, а многие из таких тэгов, содержащих «информацию об информации» так и не стали стандартом. Решение множества таких вот маленьких проблем и есть Semantic Web.

На данный момент вряд ли кто-то из людей, понимающих принципы работы семантической сети, сомневается в преимуществах этого решения или в его перспективах, но, к сожалению, сейчас не существует завершённой технологии разработки сервисов в рамках концепции семантической паутины. Перед разработчиками встаёт вопрос, изобретать что-то самостоятельно, или дождаться зрелости концепции и появления завершённых инструментов, а перед педагогами, обучать этому или нет, но чаще – как именно обучать.

Вопросы, мучающие разработчиков, в данной статье затронуты не будут, а вот попытка дать ответ на вопросы педагогов приведена ниже.

Семантическая паутина -- технология молодая, не имеющая достаточного количества примеров успешной реализации, но существующие примеры заслуживают внимания. Чего стоит один только проект Goggle Apps [<https://www.google.com/a/>] или проект Wiki [<http://wiki.org/>].

Технология снабжения всех данных метаданными, упрощающими машинный анализ, безусловно, актуальна и востребована. Сейчас семантические сервисы предлагаются в основном крупными компаниями, более того, структура сервиса – интеллектуальная собственность этой компании и делиться ей никто не будет, но это только сейчас. В будущем, когда технология шагнёт в массы, рынку понадобятся специалисты не только способные пользоваться семантическими сервисами, но и умеющие их создавать. На мысль о том, что технология шагнёт в массы, подталкивает и тот успех у пользователей, который сопровождает нынешние inside разработки крупных участников рынка, такие как сервисы Wiki или Google, и доходы этих крупных игроков.

Рынок поставщиков программного обеспечения, откликается на перспективные потребности так же быстро, как и рынок поставщиков сервисов. Все без исключения игроки на рынке специализированного программного обеспечения, предоставили инструменты разработки на языках, рекомендованных консорциумом W3 для реализации семантических сервисов. Это и технология Silverlight с собственной

средой разработки от Microsoft, и технология AJAX, используемая и поддерживаемая сообществом, и многие другие технологии и среды разработки от компаний уровня Adobe и IBM. То есть, семантическая паутина на столько интересна потенциальным пользователям, что количество разработчиков, готовых к реализации сервисов так велико, что крупные поставщики программных продуктов готовы предоставить нужные инструменты для разработки.

Из всего вышесказанного можно сделать вывод, что изучать семантические технологии и обучать им надо. То есть, ответ на вопрос, надо ли изучать и обучать, утвердительный!

Что касается процесса изучения и обучения, то всё не так однозначно. Идея семантической паутины молода, и рекомендации по реализации сервисов всё время меняются. Стандарты находятся в постоянном движении, и это порождает массу проблем. Основной проблемой является то, что авторы, использующие принципы семантической паутины или плохо представляют концепцию сети, или же не применяют идеи «искусственного интеллекта» что выливается в несколько нестандартное использование основных инструментов или применение идей сети не к документам в сети Интернет, а к чему-то научному, далёкому и не всегда для этого предназначенному, а отсутствие строгого стандарта не способствует улучшению ситуации.

На данный момент, самый распространённый вариант использования идей семантической паутины – это такой способ описания данных метаданными, который позволил бы машине использовать логику силлогизмов. Он относительно прост в реализации, ведь силлогизм знаком человеку со времён Аристотеля, и если следовать минимальным требованиям, таким как равноценность понятий, отсутствие софизмов и пр., то и выгоды от использования очевидны, и техническая реализация достаточно проста. А раз для реализации используется язык eXtensible Markup Language (XML), то любой семантический сервис сможет правильно проанализировать документ, ведь язык формализован, и структура документа полностью описана. Вроде бы одни плюсы, но...

Хотя силлогизмы и окружают нас со времен античных философов, они достигли своего апофеоза в 19-ом столетии в работах Чарльза Доджсона (более известного как Льюис Кэрролл). Доджсон написал две

книги о силлогизмах и методах их представления в графической форме, и его силлогизмы часто принимали форму сорита, при котором вывод из одной пары связанных утверждений становился новым утверждением, пригодным для связи с другими высказываниями [2].

Один из соритов Доджсона говорит:

Англичане — мужественный народ.

Мужественный народ свободен.

Свободный народ счастлив.

Следовательно, англичане счастливы

Это наводит на мысль о том, что не должным образом сформированный документ может ввести машину в заблуждение, плодом которого станет неадекватное действие, будь то просто неверный ответ поисковой машины или хищение пользовательских данных.

То есть, несмотря на привлекательную простоту, силлогизмы не слишком хорошо работают в реальном мире, потому что большинство данных, которыми мы пользуемся, не поддаются таким простым способам обработки. Как результат, семантическая сеть в нынешнем её состоянии не сможет быть однозначно полезной.

Авторы концепции семантической паутины существенно переоценивают значение дедуктивного способа вывода знаний (это часто встречающаяся проблема в проектах искусственного интеллекта). Величайшим популяризатором этой ошибки был Артур Конан Дойл, чьи истории о Шерлоке Холмсе нанесли больший урон нашему пониманию человеческого разума, нежели что-либо еще. Дойл убедил поколения читателей, в том, что, в момент размышления действительно умные люди, приходят к неизбежным заключениям посредством связывания последовательных фактов. Как Холмс лихо утверждал, «когда вы устраните невозможное, то оставшееся и будет истиной, как бы неправдоподобно это не выглядело».

Это чувство так притягательно именно потому, что оно описывает мир гораздо более простой, нежели реальный. В реальном же мире мы обычно оперируем неполной, неубедительной или чувствительной к контексту информацией. Когда нам приходится принимать решение, основываясь на этой информации, мы пытаемся строить догадки, экстраполировать, обращаемся к интуиции, поступаем так, как мы

поступали раньше в аналогичной ситуации, делаем то, что, как нам кажется, сделали бы наши друзья, или Иисус, или Иван Грозный. Мы делаем все эти вещи, и многое другое, но почти никогда, не обращаемся к дедуктивной логике.

И как следствие, почти ни одно утверждение, из тех, что мы делаем, даже самые очевидные, не является истиной в том смысле, как этого требует семантическая сеть. Дрю МакДермотт в его блестящей «Критике чистого разума» [1] взял под сомнение, что возможно создать искусственный интеллект, построив достаточно детализированную дедуктивную машину. Он заключил, что этот подход фатально недоработан, и заметил, что: - «Это, должно быть случай, когда нам просто хочется, чтобы значительная часть логического вывода была дедукцией, в противном случае это просто не будет соответствовать количеству теорем, дедуктивно выведенных из заданного набора аксиом». Хотя «Критика чистого разума» предворяла не только семантическую сеть, но и «обычный» WWW, критика все еще актуальна.

Из вышесказанного трудно сделать однозначные выводы относительно того не только как обучать, но и собственно, чему. Инструментов великое множество. Единообразие языков разработки только внешнее, концепции нет, практическая польза неоднозначна. Но, ведь кому-то это нужно, раз существуют SilverLight и NetBeans. Более того, это нужно не просто кому-то, а таким корпорациям как IBM и Microsoft, а раз интерес имеют такие крупные игроки, то и концепция будет вылизана, и варианты практической реализации проработаны, и положительные потребительские качества станут очевидными.

Например, для чего-то же существует концепция иерархии и таксономии? Это самая естественная классификация. Все остальные можно автоматически вывести из нее, исходя из свойств и отношений.

Дескриптивные логики, исходят из того, что все возможно, пока не доказано обратное. Это не закрытый мир и силлогизмы нужны, в первую очередь не для моделирования мышления, а для дополнительной связи отношений и онтологий, то есть, выстраивания сети семантики для получения адекватных запросу результатов. К тому же, возникает ощущение того, что нынешние разработчики (да и авторы) помнят логику на уровне высказываний, а про другие логики забыл. Для борьбы с

субъективностью информации давно существуют свои технологии, например, модальная логика, нечеткая логика, факторы уверенности. Безусловно, здесь есть свои проблемы, но, тем не менее, всё это можно будет применить.

Представления о мире отличаются у всех, но это не мешает успешному взаимодействию на том или ином уровне. А на каком, зависит от целей и общности интересов. Так пусть будет больше онтологий, больше разных проекций. Безусловно, на данный момент, технология семантической сети - это не для всех. Для корпораций, ассоциаций, сообществ и прочих самое то, а для остальных – пусть пока будет Web, пусть подождут - время покажет.

Всё это ещё больше запутывает, но только на первый взгляд. Из вышесказанного становится ясно, что педагог, должен уметь видеть проблемные моменты новых технологий и при этом уметь объективно оценить их перспективность. Ведь проблемы рано или поздно разрешатся, а актуальность останется.

Кроме этого, педагог должен уметь вычленять нечто общее, основополагающее в новой перспективной технологии, нечто такое, что не претерпит кардинальных изменений в процессе становления технологии и не потеряет практической значимости. В концепции семантической сети этой основой являются в равной части логика и язык XML. Язык XML – базовый язык разработки WEB приложений настоящего и будущего, перечислять его преимущества перед такими языками как Hypertext Markup Language (HTML) – тема отдельной статьи, суть в том, что они бесспорны. Все сопутствующие языки, для реализации семантических сервисов, такие как язык описания ссылок, язык таблиц стилей, язык шаблона документов, язык для описания баз данных, и многие другие, имеют XML синтаксис, что приводит всю разработку к одному общему знаменателю и упрощает труд разработчиков. Логика – отдельная тема. Думаю, тот факт, что гибко применять ту или иную логику в зависимости от задачи непременно пригодится и при разработке приложений и в повседневной жизни, ни кем оспариваться не будет.

В заключении, хочется ещё раз акцентировать внимание на основных идеях, изложенных в этой статье. Сети семантической, то есть такой, в которой машина сможет анализировать информацию, быть! На данный

момент речь не идёт об осмыслении информации, самосознании машины или другой фантастике. Речь идёт о том, что машина сможет однозначно отличить электронную визитную карточку партнёра по бизнесу от порнографической картинки или, сгруппировать информацию из нескольких источников на одной странице по тем или иным принципам (яркий пример этого – технология RSS).

Какой будет семантическая сеть, такой, как видят её сейчас, имеющей несколько полярных концепций, или совсем другой – не известно, но, она будет уже совсем скоро. Технологии XML, идея описания данных метаданными и идея построения связей, собственно сети, будут воплощены в семантической сети. То есть, умение применять логику, и навыки XML программирования – это именно то, чему надо обучать, а главное, нет никакой проблемы с тем, как это делать, более того, уже есть опыт.

Литература

1. McDermott D. Critique of Pure Reason / Drew McDermott // Computational Intelligence. - 1987. - Vol. 3, Issue 1. - P. 151-160.
2. Gardner M. 1996. The universe in a handkerchief: Lewis Carroll's mathematical recreations, games, puzzles, and word plays / Martin Gardner. New York : Copernicus, 1996. - 158 p.
3. Weiser M. Some Computer Science Issues in Ubiquitous Computing / Mark Weiser // CACM. - 1993. - Vol. 36, Issue 7. - P. 74-83.